

Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with *ANTsR*

Nicholas J. Tustison^{a,1}, K. L. Shrimidhi^b, Max Wintermark^a, Christopher R. Durst^a, James C. Gee^a, Murray C. Grossman^c, Brian B. Avants^b

^a*Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA*

^b*Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia, PA*

^c*Department of Neurology, University of Pennsylvania, Philadelphia, PA*

Abstract

Automated brain tumor segmentation in MRI is complicated by the lack of prior knowledge concerning tumor location, spatial extent, shape, possible displacement of normal tissue, and intensity signature. These factors limit the applicability of techniques with established performance in related tasks such as normal brain tissue segmentation. We introduce a template-based framework for computing geometric and asymmetry-driven feature sets from multiple modality neuroimaging. We use these features to drive a supervised whole-brain and tumor segmentation approach based on random forests. The asymmetry-related features demonstrate excellent discriminative properties within this framework. This approach is made possible by symmetric multimodality brain templates which are optimal in terms of both shape and multivariate intensity. Further performance improvements are demonstrated by a

¹Corresponding author: PO Box 801339, Charlottesville, VA 22908; T: 434-924-7730; email address: ntustison@virginia.edu.

successive application of random forest models for increasing levels of feature refinement to which we refer as *concatenated random forests*. Although the derived workflow involves many disparate components, the newly developed *ANTsR* open source software package significantly facilitates the proposed methodology. *ANTsR* is a comprehensive interface between the popular Advanced Normalization Tools (ANTs) and the *R* statistical project. *ANTsR* integrates proven image analysis methods for image registration, segmentation, bias correction, and template building with *R*'s statistical and machine learning techniques and state-of-the-art visualization possibilities. The reported algorithmic framework was the top-performing entry in the MICCAI 2013 Multimodal Brain Tumor Segmentation challenge and, to our knowledge, is the only entry for both challenge years (2012 & 2013) which has been made publicly available.

Keywords: advanced normalization tools, BRATS, brain tumor segmentation, *R* project for statistical computing

1. Introduction

Given the complexity of tumor growth and appearance and the need for precise volumetric measurements for tumor characterization, much research has been invested in computational approaches for automated segmentation of tumor regions in MR images. Several approaches have been previously proposed in the literature as detailed in recent reviews (Angelini et al., 2007; Bauer et al., 2013). Commensurable algorithmic evaluation, however, is extremely problematic due to widely varying performance assessments described in the corresponding publications, lack of publicly available evaluation data, and private algorithmic instantiations. In response to these issues, the Multimodal Brain Tumor Segmentation (BRATS) challenge was initiated in 2012 (and continued in 2013) under the auspices of the Medical Image Computing and Computer Assisted Intervention Society in association with their annual international conference (Menze et al., 2014).

After observing the 2012 challenge to gain insight for our own research needs (Durst et al., 2014), we found that no teams had availed their software in any form. This prompted our participation in the challenge which took place the following year in Nagoya, Japan. Heavily influenced by the success of participants in the 2012 challenge (specifically, the work of Bauer et al. (2012); Geremia et al. (2012); Zikic et al. (2012)), we adopted the increasingly popular random forest (RF) machine learning framework (Breiman, 2001) which permits the inclusion of many potentially discriminative image features. Supervised segmentation techniques generally consist of a training phase for model construction using image-based feature data followed by prediction using the generated model. For supervised brain tumor segmen-

tation, a set of training data consisting of labeled brain images is used to construct a predictive model. Although other classification techniques have been used to segment brain tumors (e.g., support vector machines (Bauer et al., 2011)), RF models have proven particularly successful.

Several machine learning concepts were integrated to create the RF framework first articulated in its entirety by Breiman et al. (Breiman, 2001) for performing classification/regression. Although decision trees had been previously explored in the literature, it was the success of “boosting”-style machine learning techniques, such as AdaBoost (Schapire, 1990; Freund and Schapire, 1997), which influenced the aggregation of such decision trees into “forests” with randomized node optimization for improved classification/regression performance (Ho, 1995; Amit and Geman, 1997). The final element of bootstrap aggregating or “bagging” (i.e. random sampling of the training data) was introduced by Breiman (Breiman, 1996) to achieve improved accuracy.²

Early adoption Viola et al. (2005) and success in the computer vision community has led to a recent surge within the medical image analysis community of using RFs for handling complex classification/regression tasks including normal brain segmentation (Yi et al., 2009), MS lesion segmentation Geremia et al. (2011), multimodal brain tumor segmentationp (Bauer et al., 2012; Zikic et al., 2012), brain extraction (Iglesias et al., 2010), anatomy detection in computed tomography (Criminisi et al., 2013), and segmentation

²One of the principal advantages of *R* is the extensive community of developers who have contributed on the order of thousands of packages extending *R*’s capabilities beyond its core functionality. Most relevant here is the **randomForest** package developed from Breiman’s original Fortran code by Liaw and Wiener (Liaw and Wiener, 2002).

of echocardiographic images (Verhoek et al., 2011). A thorough introduction for those interested in delving deeper into the more theoretical aspects of RFs can be found in Criminisi et al. (2011).

Instrumental to our success in the 2013 challenge was the use of image features based on multimodal shape and appearance asymmetries which were generated via the construction of symmetric, multivariate brain templates. Additional improvements in performance were due to the use of a two-stage RF model approach whereby the output probability images from the application of the first RF model seeds the construction of a subset of new feature images for input into the second stage. The results of the second RF model are then refined using a series of heuristically-derived binary morphological operations. The two-stage approach is motivated by the fact that the RF is a powerful tool for ranking the features in a high-dimensional model i.e. can be used for model selection. However, the RF outputs are not geometrically constrained, which is the motivation for our second refinement step.

In line with our philosophy regarding open science (Tustison et al., 2013; Ince et al., 2012), we have made the code available as open source³ and, to our knowledge, we are the only competitors out of the approximately 20 teams that participated in either years' challenges to do this. All code is based on the well-vetted and open source Insight Toolkit (ITK)⁴ of the National Institutes of Health and is available through our *ANTsR* package. In subsequent sections we describe in greater detail the automated supervised brain segmentation pipeline including the generation of asymmetry feature

³<https://github.com/ntustison/BRATS2013>

⁴<http://www.itk.org/>

images and the concatenated RF construction. We also describe the code and data used for the BRATS challenge so that the interested reader can reproduce our results. Finally, we illustrate performance with results from the BRATS 2013 challenge.

2. Materials and Methods

Certain key elements characterize the proposed supervised brain tumor segmentation protocol including:

- construction of symmetric multivariate templates for multimodal asymmetry-based features,
- generation of other image-based features,
- training of RF models, and
- prediction using the proposed framework.

In subsequent sections, we describe these items and detail the *ANTsR* framework which coordinates all steps.

2.1. Image Features for RF Model Training and Prediction

2.1.1. Multimodal Asymmetry Features from Multivariate Symmetric Templates

In order to better characterize deviations from normal brain shape and appearance, several image features were derived using symmetric population-specific multivariate templates. For normal neuroanatomy, the use of spatial prior information coupled with image normalization capabilities has proven

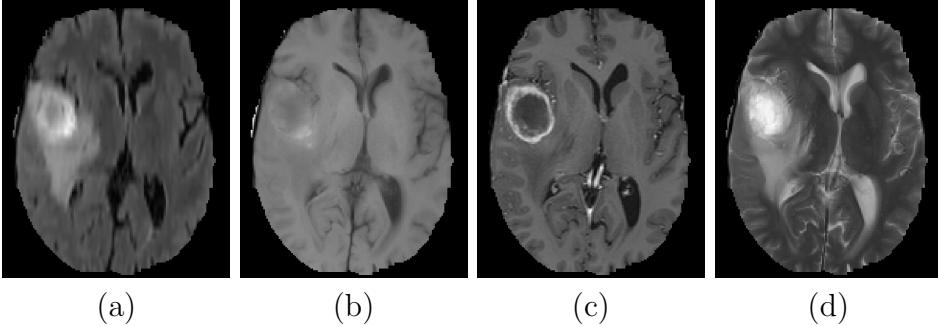


Figure 1: Induced bilateral asymmetry due to tumor presence causing distortion of the plane of symmetry. Shown are mid-axial slices of one of the BRATS 2013 training data (specifically `BRATS_HG0001`): (a) FLAIR, (b) T1, (c) T1C, and (d) T2.

useful in producing improved segmentation results of “expected” brain tissue such as cerebrospinal fluid, gray matter, and white matter (e.g., Ashburner and Friston, 1997). In contrast, accommodating spatial priors to model the presence of a possible tumor and its constituent tissue components is difficult. However, since the normal brain exhibits a bilaterally symmetric organization, we can use the presence of asymmetries to potentially differentiate abnormal brain tissue. A similar motivation prompted the identification of the mid-sagittal plane of symmetry (Prima et al., 2002) for feature generation in multiple sclerosis lesion (Geremia et al., 2011) and tumor (Geremia et al., 2012) identification. However, this earlier approach does not take into account the displacement of normal tissue due to tumor growth causing the mid-sagittal plane to deform from its planar structure (cf Figure 1).

To take into account these potential asymmetries, we require a data set with the same modalities as dictated by the subject image acquisition protocol. Although it would be preferable to build population-specific multivariate templates from normal data using the same acquisition parameters (Avants et al., 2010), we substituted well-known, publicly available data since such

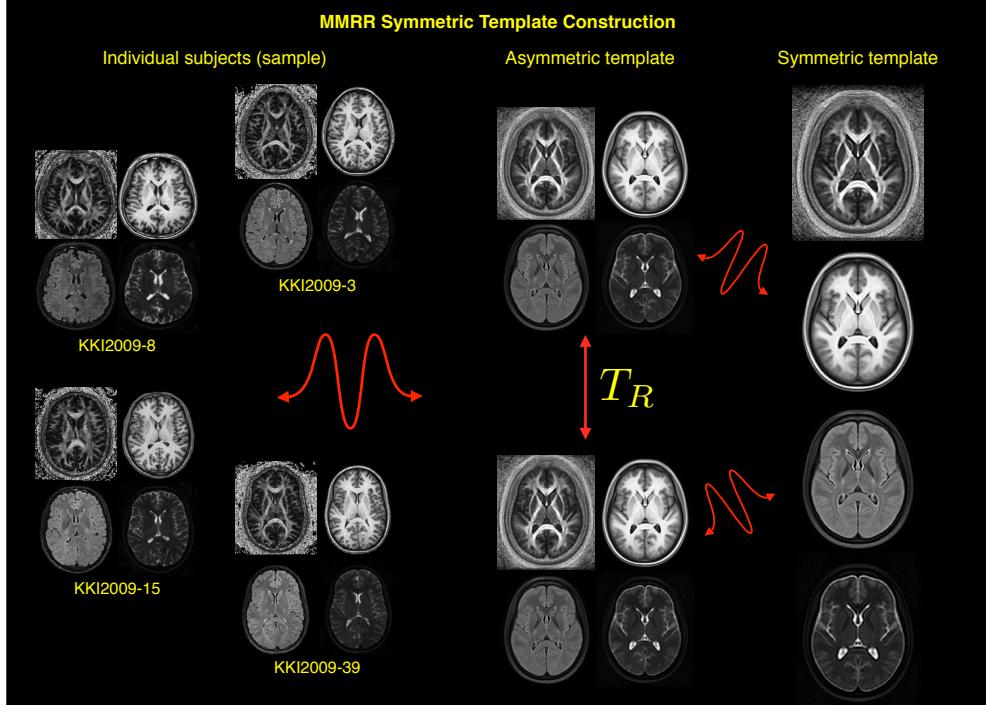


Figure 2: Multivariate symmetric template created from the MMRR data (Landman et al., 2011). Of the seven modalities comprising the set of study acquisitions, we illustrate the (a) FA, (b) FLAIR, (c) MPRAGE, and (d) T2 template components. Although DWI-based images were not included in the challenge data, such images have shown discriminative potential (Price et al., 2003; Cha, 2005) warranting investigation in our future work. The optimal transformation and averaging of the individual subject images result in the asymmetric template represented at the top of the middle column. A horizontal reflection perpendicular to the mid-sagittal plane, T_R , resulted in the contralateral counterpart represented at the bottom of the second column. The final template seen on the right is a result of repeating the template construction using the two asymmetric templates as input.

normal data was not made available for the BRATS 2013 challenge. A recent neuroimaging reproducibility study by Landman et al. resulted in an open data cohort of 21 normal individuals, each imaged twice, comprising several modalities including ASL, FLAIR, DTI, fMRI, T1, and T2 (Landman et al., 2011). These data (known as the “MMRR” data set) were selected for deriving a multivariate template due to its public availability and inclusion of several modalities (even permitting future incorporation of modalities not currently included with the BRATS challenge into our segmentation framework).

As detailed in Avants et al. (2008a, 2010), given K multimodality images, $\mathbf{I} = \{I_1, I_2, \dots, I_K\}$, for N subjects, multivariate template construction iterates between optimizing the set of diffeomorphic transforms between the subjects and the template, $\{(\phi_1, \phi_1^{-1}), \dots, (\phi_N, \phi_N^{-1})\}$ and constructing the optimal multivariate template appearance $\mathbf{J} = \{J_1, J_2, \dots, J_K\}$ (with corresponding coordinate system $\psi(\mathbf{x})$) to minimize the following cost function:

$$\sum_{n=1}^N \left[D(\psi(\mathbf{x}), \phi_1^n(\mathbf{x}, 1)) + \sum_{k=1}^K \lambda_k \Pi_k(I_k^n, J_k(\phi_n^{-1}(\mathbf{x}, 1))) \right] \quad (1)$$

D is the diffeomorphic shape distance,

$$D(\phi(\mathbf{x}, 0), \phi(\mathbf{x}, 1)) = \int_0^1 \|\nu(\mathbf{x}, t)\|_L dt \quad (2)$$

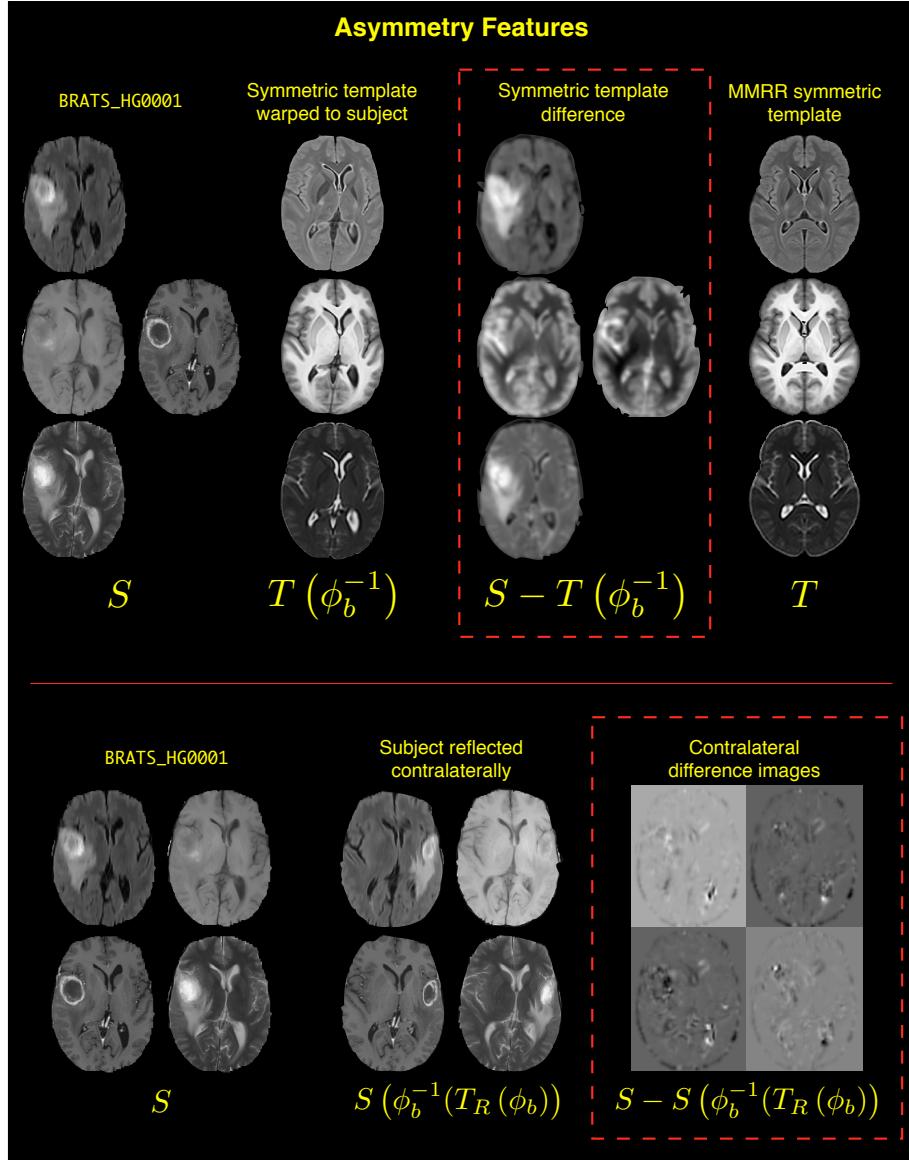


Figure 3: Given the mapping between the template, T , and subject, S , domains ($\phi_b : S \xleftrightarrow{b} T$), various features can be calculated which demonstrate good discriminative qualities. Feature images used are outlined in a dashed line. Top: Difference images with the symmetric multivariate template are created by warping the template to the subject space and performing a voxelwise subtraction from the original modality image. Bottom: Similarly, contralateral difference images are calculated from each modality per subject by generating the non-Euclidean contralateral image via the diffeomorphic transform ϕ_b .

dependent on the choice of linear operator, L , and ν is the velocity field

$$\nu(\phi(\mathbf{x}, t)) = \frac{d\phi(\mathbf{x}, t)}{dt}, \quad \phi(\mathbf{x}, 0) = \mathbf{x}. \quad (3)$$

Each pairwise registration employing the similarity metric Π_k can be assigned a relative weighting, λ_k , to weight a particular modality's influence in the construction process. Once the multivariate template has converged (typically in four iterations), we symmetrize the template by flipping each asymmetric template component contralaterally and then running the multivariate template construction a second time using only the multivariate template and its symmetric analog. This is illustrated conceptually in Figure 2.

After constructing the template offline, each data set is processed by first registering the non-contrast T1-weighted image to the T1-weighted component of the symmetric template. To do this we use a recently developed SyN (Avants et al., 2011a) variant based on B-spline regularization which has demonstrated good performance in normal brain registration (Tustison and Avants, 2013). Good performance also extends to this pathological data scenario as indicated by visual inspection and the fact that the derived features were amongst the most informative in our winning entry.

We denote the mapping from the subject, S , to the template, T , space as $\phi_b : S \xleftrightarrow[b]{\sim} T$ which consists of both affine and diffeomorphic components. Transform invertibility is essential for the template-based features. The first set of feature images is generated by warping the template components to the subject space and calculating the difference image (see the top portion of Figure 3). For example, the T2 symmetric template voxelwise difference

image is calculated from

$$\text{T2 symmetric template difference} = S_{\text{T2}} - T_{\text{T2}}(\phi_b^{-1}). \quad (4)$$

Note that the T1-weighted contrast and non-contrast images are paired with the T1-weighted component of the symmetric template. The second set of template-based feature images is generated per modality as the difference image with the contralateral reflection. This is achieved by calculating the reflection transform, T_R , in the symmetric template space and composing transforms as follows to create the non-Euclidean contralateral counterpart per modality:

$$S_{\text{contralateral}} = S(\phi_b^{-1}(T_R(\phi_b))). \quad (5)$$

This process is illustrated in the bottom portion of Figure 3. A final, related feature is the Jacobian of the transform. The basis for this final feature is that relatively larger Jacobian values might be indicative of tumor expansion.

2.1.2. Image Features for Concatenated Random Forests

Our segmentation protocol involves training and application of two RF models in succession (see Figure 5) to which we refer as “stages.” The basic idea is that we generate a set of feature images used as input to the first RF model (or first stage) which produces a voxelwise probabilistic tissue estimate⁵ for the seven brain/tumor tissue types:

⁵More precisely, in the RF framework, each prediction sample (i.e. the feature vector at each voxel), is propagated through each tree of the ensemble where it is labeled as belonging to a specific class. These “votes” are converted to voxelwise probabilistic estimates for

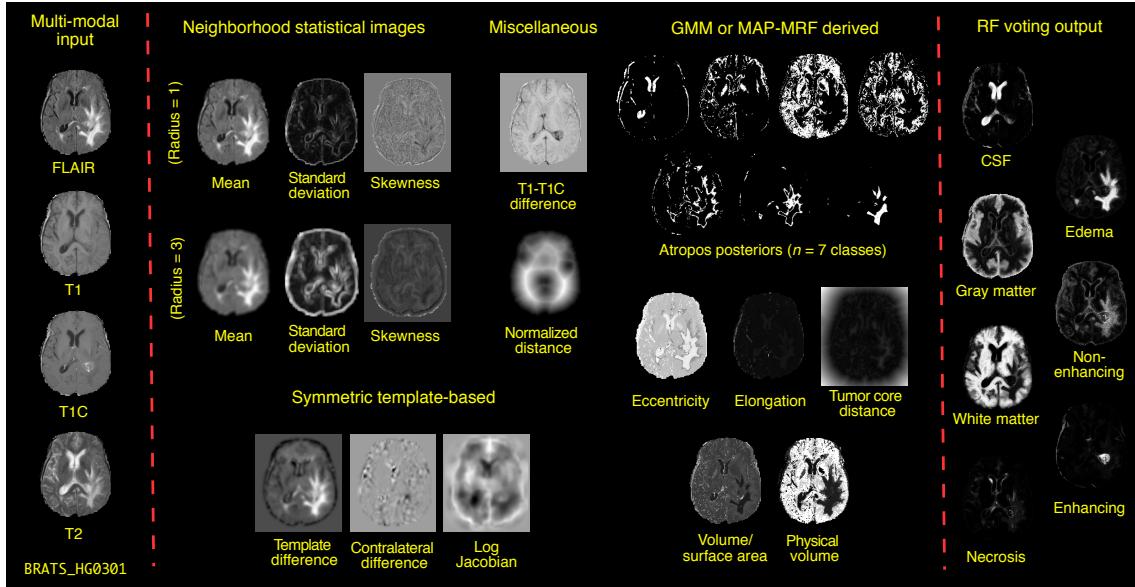


Figure 4: Representative feature images derived from the BRATS_HG0301 “challenge” data set. Neighborhood statistical images for each modality were generated by calculating a given statistic within a specified neighborhood radius. Also calculated for each modality were feature images based on either the GMM or the MAP-MRF segmentation. For the former, we show the probability maps for each of the seven labels which are used as feature images. From the resulting hard segmentation, we calculate various geometric measures per connected component of each of the seven labels. Similarly, the registration to the symmetric template produces the modality-specific difference images with the corresponding symmetric template itself and with respect to the contralateral side. This mapping is also used to produce the log Jacobian image. Finally, the ($T_1 - T_{1C}$) image is calculated and, from the cerebral mask, we calculate the normalized distance image.

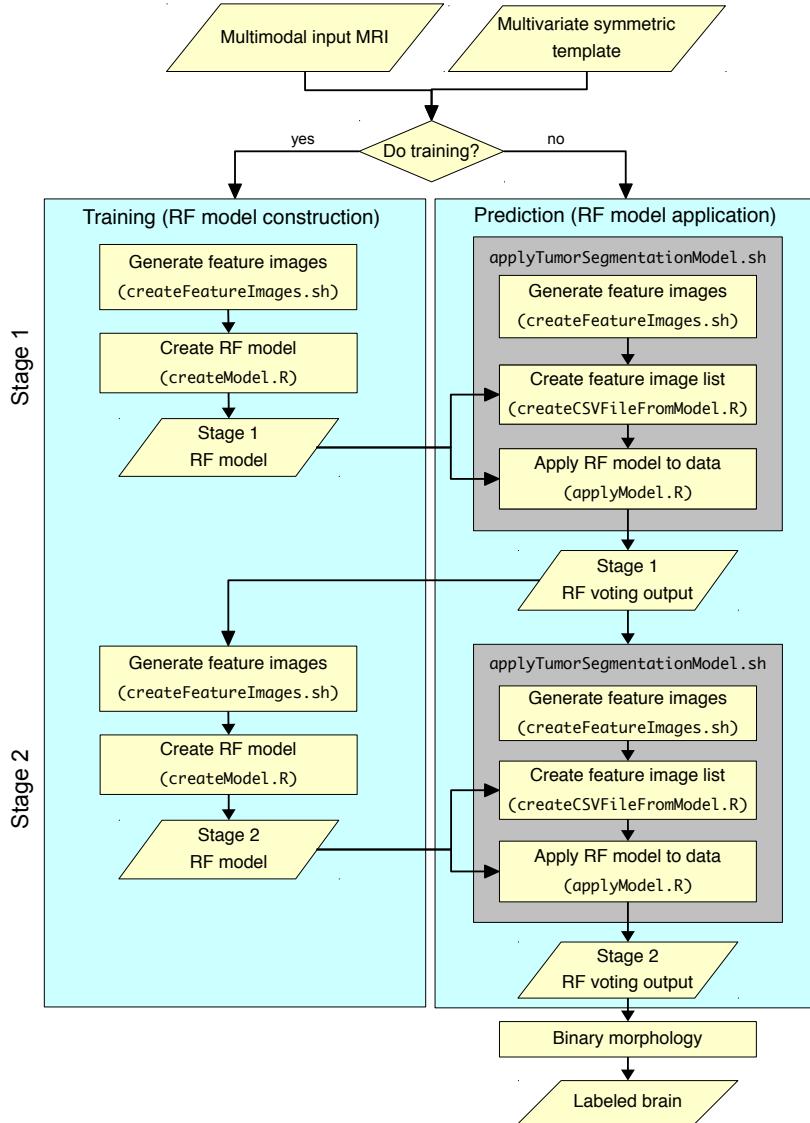


Figure 5: Diagrammatic workflow for the proposed RF model training and prediction. The feature images are first generated from the input MRI and symmetric multivariate template. If training, the set of feature images derived from the training data are used to create the RF model for the first stage. The model is then applied to the training or prediction data to generate a second set of feature images to create the second stage RF model (if training) and/or generate the second RF model. For prediction, the final refinement process entails a series of heuristically-derived binary morphological resulting in the final labeled brain.

- cerebrospinal fluid,
- grey matter,
- white matter,
- edema,
- non-enhancing tumor (including low-grade tumor center),
- enhancing tumor (excluding necrotic center), and
- abnormal necrotic center or necrocyst in high-grade gliomas.

We then use these output tissue map estimates as spatial priors for generating a second set of image features. These are used as input for a second RF model application (or second stage), the output of which constitute the final tumor segmentation estimate. Note that some feature images are included in both stages such as the asymmetry features described in the previous section.

For both stages, in contrast to previous generative modeling approaches for multimodal tumor segmentation (e.g., Prastawa et al., 2003), we do not use multivariate Gaussians to specify tissue probabilities but rather incorporate each univariate probability map into the feature vector of the training data. As pointed out in Menze et al. (2010), multivariate modeling might obscure the distinct biological information provided by each modality. Instead, we let the RF construction process determine the optimal combination of such multivariate information. Additionally, maximum posterior labeling

each class.

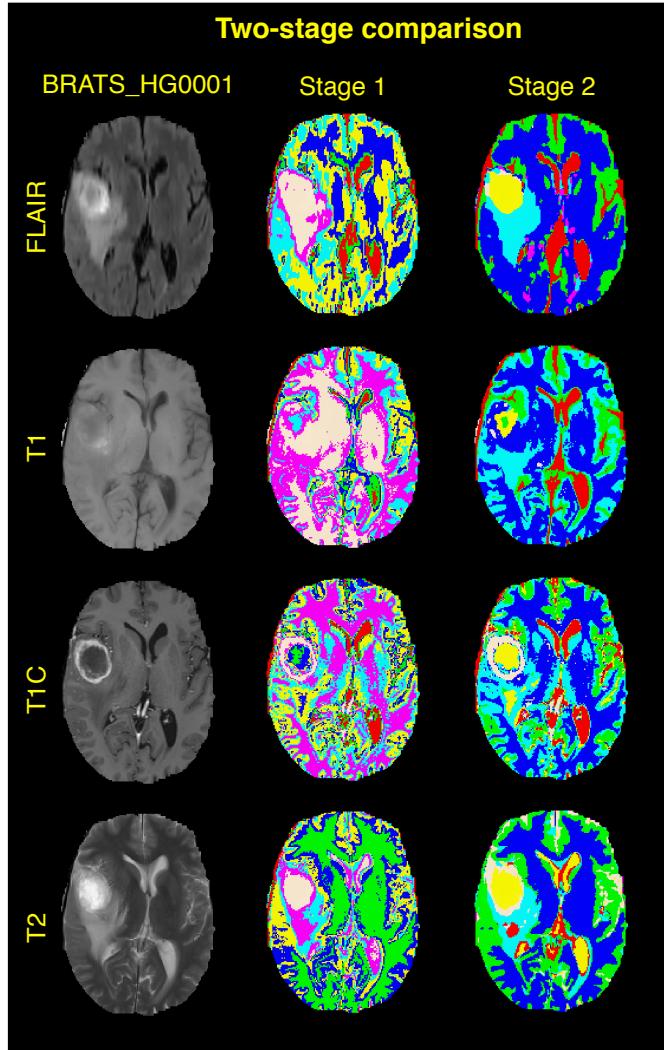


Figure 6: Visual comparison between the first and second RF stages on the evaluation BRATS_HG0001 data set. Although the GMM intensity modeling does provide certain discriminative benefit (see Figures 8 and 9), the use of the RF-derived posteriors coupled with the described MAP-MRF segmentation framework enhances the accuracy of the intensity-based modeling features.

from both stages is used to determine the connected components for each label. Geometric features (assigned voxelwise) include the physical volumes of each connected component, the volume to surface area ratio, the elongation, and eccentricity.

Stage 1 (Initial voxelwise classification). Gaussian mixture modeling (GMM) is used to assign voxelwise tissue probabilities which comprise an additional set of image features where prior cluster centers for specific tissue types for each modality are learned from training data (Reynolds, 2009). Intensity normalization across the training cohort was originally performed using Nyúl et al. (2000). However, despite good performance on normal data, application of this intensity normalization algorithm (after a thorough parameter search) on BRATS data led to slightly worse performance than simply winsorizing the intensity values to the quantile range [0.01, 0.99] and then rescaling the resulting intensities to [0, 1] so we opted for the latter, more simplistic intensity normalization protocol. The cluster centers are defined as the mean normalized intensity value for each tissue type of each modality image over all the training data. It should be noted that initializing with these values is optional and that results seem to be robust to initialization values.

A GMM formulates the probability distribution at each voxel, \mathbf{x} , as the sum of M Gaussian components, $\mathcal{N}(\mathbf{x}|\mu, \sigma)$, i.e.

$$p(\mathbf{x}|\mu_m, \sigma_m, \lambda_m) = \sum_{i=1}^M \lambda_m \mathcal{N}(\mathbf{x}|\mu_m, \sigma_m) \quad (6)$$

where $\sum_{m=1}^M \lambda_m = 1$. The parameters of the GMM are determined with maximum likelihood estimation using the Atropos segmentation tool (Avants

et al., 2011b) available in ANTs. We include the resulting seven posterior probability tissue images along with five additional geometric features based on the connected components of the segmentation image (distance image to tumor core, physical volume image, volume to surface area ratio, eccentricity, and elongation). During the training phase (or RF model construction), these feature images, $\{F_i : i \in GMM\}$, generated during Stage 1 are used to create the Stage 1 RF model using the following relationship

$$\text{Tissue classes} \sim \sum_{i \in GMM} F_i + \sum_{i \in Asym} F_i + \sum_{i \in Misc} F_i \quad (7)$$

given in the standard R notation of Wilkinson and Rogers (1973). The set of $\{F_i : i \in Asym\}$ are the feature images generated from the symmetric multivariate template described in the previous section and $\{F_i : i \in Misc\}$ are miscellaneous feature images described in Section 2.1.3.

Stage 2 (Refined spatial coherence classification). Application of the Stage 1 RF model to a set of feature images from an individual subjects results in a set of images estimating the spatial probabilities for each of the seven tissue types. At each voxel, the corresponding multi-feature vector is propagated through each tree of the random forest resulting in a single vote/classification per tree. The classification probabilities are derived by normalizing these votes to $[0, 1]$.

At the beginning of the second stage, these spatial probability maps are used as spatial priors in a MAP-MRF segmentation also performed using the Atropos segmentation tool (Avants et al., 2011b) available in ANTs although we couple it with N4 bias correction (Tustison et al., 2010) as contained in the

script `antsAtroposN4.sh`. The probabilistic segmentation output and corresponding connected-component feature images for each modality replace the analogous GMM-based feature images used during the first stage of training/prediction along with the other feature images generated during the feature stage (such as the template-based data). The second stage RF model is then produced similarly as the previous stage:

$$\text{Tissue classes} \sim \sum_{i \in MAP-MRF} F_i + \sum_{i \in Asym} F_i + \sum_{i \in Misc} F_i. \quad (8)$$

The use of these MRF features improves the accuracy of the intensity-based modeling over the GMM approach by introducing a spatial coherence constraint to those features in addition to providing a better initialization with the RF posteriors from the initial estimation stage.

2.1.3. Miscellaneous Multimodal Feature Image Generation

Our strategy for feature selection was to, first, generate an ample feature set and, second, to use the training phase to prioritize the features appropriately. In principal, this will eliminate weak features and retain strong ones even in the presence of substantial noise (cite some RF paper here ...). For each modality, several images were generated based on calculation of voxel-wise neighborhood first-order statistics including mean, variance, skewness, and entropy using neighborhood block radii of 1 and 3 voxels. We also calculated two Euclidean distance transforms (Maurer et al., 2003) to be used for each subject. One was calculated from the subject's own cerebral mask and the other was the distance transform of the symmetric template cerebral mask warped to the subject. Finally, we also generated the (T1 - T1C) dif-

ference image (Prastawa et al., 2003). The script `createFeatureImages.sh` performs all preprocessing and generates all feature images. In Figure 4 we provide sample mid-axial image slices from features generated from the `BRATS_HG0301` data set.

2.2. Overview

In summary, the proposed RF classification approach per subject consists of the following steps:

1. Calculate stage 1 features based on:
 - GMM intensity modeling,
 - asymmetry, and
 - miscellaneous.
2. Apply stage 1 RF model to the above calculated features resulting in seven probability maps for the seven tissue types.
3. Use the probability maps to calculate stage 2 features based on MAP-MRF intensity modeling.
4. Apply stage 2 RF model to MAP-MRF, asymmetry, and miscellaneous features to get final RF classification.

2.3. ANTsR: An ANTs/R Interface

The complexity of neuroimaging research necessitates commensurable numerical analysis capabilities. Similarly, concomitant with the era of “big data” (specifically with respect to neuroimaging (Van Horn and Toga, 2013)) are new visualization needs and challenges (Childs et al., 2013; Kehrer and Hauser, 2013). In response, various software packages have been developed

to integrate tools specific to neuroimaging research with more general numerical and visualization software packages. The well-known neuroimaging package SPM⁶ is a significant extension of the commercial computing and visualization environment Matlab. Open source neuroimaging packages, such as NIPY (neuroimaging in Python),⁷ rely on other open source packages for numerical/statistical analysis. NIPY, for example, uses the more generic packages NumPy and SciPy for numerical analysis and optimization.⁸

ANTs (Advanced Normalization Tools) was built, originally, to provide high performance image registration for medical image analysis (Avants et al., 2008b) and based upon the mature Insight Toolkit (ITK) sponsored by the National Institutes of Health. Since then, ANTs has grown to include several robust medical image analysis solutions including bias correction (Tustison et al., 2010), *n*-tissue multivariate segmentation (Avants et al., 2011b), template construction (Avants et al., 2010), and cortical thickness estimation (Das et al., 2009) (many of which have been introduced into ITK partially in an attempted leveraging of Linus’s Law—“Given enough eyeballs, all bugs are shallow”). However, in the evolution of the toolkit, it became clear that robust statistical machinery was lacking for making inferences regarding the data produced during the course of ANTs processing. *ANTsR* was developed specifically to provide an interface between ANTs, a powerful neuroimaging toolkit for producing reliable imaging data transformations, and the *R* project⁹ for statistical computing and visualization thus providing a complete

⁶<http://www.fil.ion.ucl.ac.uk/spm/>

⁷<http://nipy.org>

⁸<http://www.numpy.org>

⁹<http://www.r-project.org>

set of tools for multivariate neuroimage analysis. *ANTsR* intends to provide a modern framework for medical analytics, with a focus on imaging-assisted prediction and statistical power.

Careful consideration of available statistical software led to the adoption of *R* to complement ANTs quantification resulting in the *ANTsR* package. *R*'s open source code base, reliable software testing and distribution strategies, and add-on packages coupled with its rapidly growing community of developers and users has caused wide-scale adoption within both academia and industry.

2.3.1. Installation

The *ANTsR* package is publicly available on the github project hosting service.¹⁰ Prior to installation of *ANTsR*, several external *R* packages need to be installed including: `Rcpp`, `signal`, `timeSeries`, `mFilter`, `doParallel`, `robust`, `magic`, `knitr`, `pixmap`, `rgl`, and `misc3d`.¹¹ Additionally, in order to perform the supervised brain segmentation as described in later sections, one needs to also install the packages `randomForest`, `snowfall`, `rlecuyer`, and `ggplot2`.¹²

In addition to *R* and the add-on packages previously mentioned, CMake is also required. CMake¹³ is an open source tool for the management and building of large-scale software projects. It is used to coordinate the downloading of external packages, such as the Insight Toolkit (ITK)¹⁴ and ANTs. Fur-

¹⁰<https://github.com/stnavar/ANTsR>

¹¹See <http://stnavar.github.io/software/2014/01/08/antsr/> for current status.

¹²Packages are easily installed using the `install.packages()` *R* mechanism.

¹³<http://www.cmake.org/>

¹⁴<http://www.itk.org/>

ther instructions for download and installation can be found on the *ANTsR* github website. Feel free to contact the authors if installation trouble occurs. We note that *ANTsR* is currently only tested on UNIX-alikes such OSX and Ubuntu operating systems.¹⁵

2.3.2. Usage

ANTsR is intended to not only allow easy interchange between medical imaging formats and *R* but also to facilitate reproducible scientific studies and the type compilable analysis articles that are fundamental to journals such as *Biostatistics*. Both **knitr** and **sweave** facilitate integration of R-code with the LaTeX document preparation system.

An additional motivation for our development of *ANTsR* (and hopefully its acceptance by the community) stems from the ability to couple ANTs core functionality, including IO tools such as **antsImageRead**, with the large number of *R* statistical and visualization packages. Due to this combination, several functions have been easily created for such neuroimaging-specific tasks as fMRI/ASL data manipulation and analysis, voxel and ROI-based analyses, and connectivity visualization. The user help menu and documentation for the library and its constituent functions are invoked in the similar manner as other R libraries.

As mentioned earlier, we have made this entire framework available as open source. In addition to the *ANTsR* repository already on github which houses both ANTs and *ANTsR* functionality, we created a special github

¹⁵Windows installation should be possible but, to our knowledge, has not been attempted.

repository specifically for this work containing figures, references, and text.¹⁶ Also, we posted all scripts (R, shell, and perl) used to coordinate the *ANTsR* processing including:

- `applyModel.R`: applies a RF model to a new feature data set from a testing subject resulting in a set of probability images (one for each label).
- `applyTumorSegmentationModel.sh`: generates the new feature image set from the testing MRI (by calling `createFeatureImages.sh`). organizes the file names in a csv file (via `createCSVFileFromModel.R`), and applies the RF model using `applyModel.R`.
- `applyTumorSegmentationModelForCohort.pl`: Coordinates tumor segmentation on the computational cluster for a given cohort.
- `createCSVFileFromModel.R`: organizes the set of feature image file names in a csv file for input into `applyModel.R`.
- `createFeatureImages.sh`: creates the set of feature images given a set of co-registered input MRI from a single subject.
- `createModel.R`: creates a RF model given the input csv file of the feature image file names for all training data and set of truth label maps.
- `plotVariableImportance.R`: produces a plot of the importance of each feature variable used in constructing the model.

¹⁶<https://github.com/ntustison/ANTsAndArboles>

We also include a fully functional 2-D example which performs both testing and training on sample challenge data.¹⁷ After pulling the repository, one can run the scripts `exampleTrain.sh` and `examplePredict.sh` to get the sample results. Output includes several overlap measures describing the performance.

2.4. Brain Tumor Data

Brain tumor image data used in this work were obtained from the NCI-MICCAI 2013 Challenge on Multimodal Brain Tumor Segmentation¹⁸ organized by K. Farahani, M. Reyes, B. Menze, E. Gerstner, J. Kirby and J. Kalpathy-Cramer. The challenge database contains fully anonymized images from the following institutions: ETH Zurich, University of Bern, University of Debrecen, and University of Utah and publicly available images from the Cancer Imaging Archive (TCIA). Both training and testing data were made freely available through the Creative Commons Attribution-NonCommercial 3.0 license.

Training data consisted of multimodal brain MRI (T1, T2, FLAIR, and post-Gadolinium T1) from 30 glioma patients (both low, $n = 10$, and high-grade, $n = 20$, and with and without resection). For each subject, the T1, T2, and FLAIR MRI were linearly registered to the post-contrast T1. Subsequently, the brains were skull-stripped and resampled to 1 mm isotropic resolution. Testing data was processed similarly and released during the course of the challenge in two sets denoted as “Leaderboard” and “Challenge”

¹⁷<https://github.com/ntustison/ANTsAndArboles/tree/master/SimpleExample>

¹⁸<http://martinos.org/qtim/miccai2013/index.html>

data. The former consisted of 21 and 4 high and low-grade tumor patients, respectively, whereas the latter comprised 10 high-grade only patients.

Manual labeling was performed in the axial plane following a detailed protocol.¹⁹ The labeling of pathology was categorized into four regions: edema, non-enhancing tumor (including low-grade tumor center), enhancing tumor (excluding necrotic center), and abnormal necrotic center or necrocyst in high-grade gliomas. Normal brain tissue was not labeled.

2.4.1. Training: RF Creation for the BRATS 2013 Challenge

For use with the Challenge and Leaderboard data, cohort-specific models (both low-grade and high-grade glioma) for both GMM and MAP-MRF stages were created using only the supplied training data. Prior to training, we segmented normal brain tissue (Avants et al., 2011b) for each data set by segmenting only the T1 image. This was only to yield a rough estimate of normal brain tissue to augment the already provided pathology labels. This resulted in seven labels for tissues described earlier i.e., csf, gray matter, white matter, necrosis, edema, enhancing, and non-enhancing tumor characterizing each brain.

Initial testing of our proposed framework was performed on the training data using a leave-one-out strategy. Once the feature images are created for each subject, the resulting images of the entire training cohort are organized in a csv file for input into the *R* script `createModel.R`. Other possible input parameters include the requested number of trees, number of samples per label, and number of threads for parallel processing. The output is an `RData`

¹⁹<http://martinos.org/qtim/miccai2013/data.html>

file describing the RF model which can be used for future predictions.

Additionally, the `randomForest` package provides measurements for determining the importance of chosen features when constructing the model. This aids in potential feature pruning or intuiting model behavior. As mentioned earlier, we provide the *R* script `plotVariableImportance.R` to render one such quantity denoted as `MeanDecreaseAccuracy`. During model construction (specifically the out-of-bag error calculation stage), the decrease in prediction accuracy with the omission of a single feature or variable is tracked and averaged. Thus, those features which have the greatest decrease in mean accuracy are considered to be the most discriminative.²⁰ In this work, we do not use these measurements for feature pruning. However, we plot them in the Results section (see Figures 8 and 9) as they demonstrate the relative importance of our selected features including that of the proposed asymmetry images.

2.4.2. Prediction: Applying the RFs for the BRATS 2013 Challenge

Once the models are created, classification of tumors in new subjects is performed as illustrated in Figure 5. From the feature images and input GMM model, a tentative set of RF voting output confidence images are produced. As described, this is used as input to the second prediction round. The final probability output images are used to produce the maximum probability labeling.

²⁰There is an issue with consistently ranking correlated predictors as described at <http://www.r-bloggers.com/random-forest-variable-importance/> since the permutation testing performed for predictive accuracy assessment assumes predictor independence. Correctives have been proposed but we ignore these issues for this particular application.

Table 1: Assessment measures from the MICCAI 2013 BRATS Challenge including Dice overlap, positive prediction value (PPV), and sensitivity.

Data Set	Rank [†]	Dice	PPV	Sensitivity
Challenge (complete)	1	0.87 (1)	0.85 (2)	0.89 (2)
Challenge (core)		0.78 (1)	0.74 (5)	0.88 (1)
Challenge (enhanced)		0.74 (1)	0.69 (4)	0.83 (1)
Leaderboard (complete)	1	0.79 (2)	0.83 (1)	0.81 (4)
Leaderboard (core)		0.65 (1)	0.70 (2)	0.73 (2)
Leaderboard (enhanced)		0.53 (3)	0.51 (3)	0.66 (2)
Evaluation (complete)	2	0.88 (2)	0.88 (3)	0.89 (3)
Evaluation (core)		0.76 (2)	0.80 (5)	0.79 (3)
Evaluation (enhanced)		0.55 (3)	0.65 (3)	0.53 (3)

[†]Indicates performance for all measures over all regions. Component rankings are given in parentheses.

A final round of binary morphological operations were heuristically designed to improve the final segmentation results such as removal of small connected components and morphological closure of certain regions. All steps are included in the script `applyTumorSegmentationModelForCohort.pl` designed for parallel subject processing on the computational cluster at the University of Virginia.²¹

3. Results

A total of four RF models were created from the 30 training data sets. Two models for Stage 1 and Stage 2 processing were generated from the 20 high-grade glioma evaluation data described earlier. Similarly, two addi-

²¹<http://www.uvacse.edu>

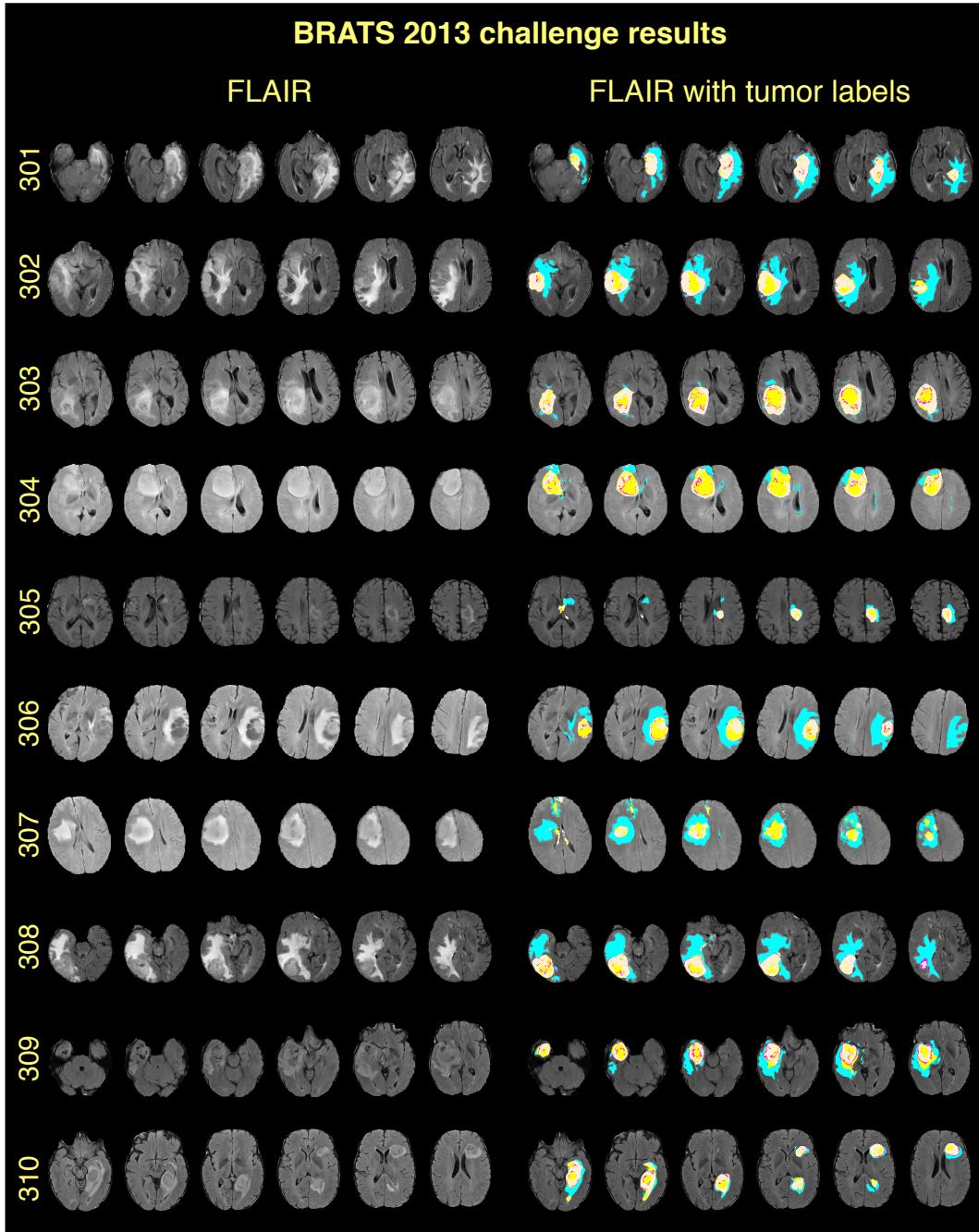


Figure 7: Visual results (every tenth slice sampled from the axial center of the tumor region) from the BRATS 2013 challenge using the proposed framework (off-white = non-enhancing tumor, cyan = edema, violet = enhancing tumor, yellow = necrosis). Tumor components were combined into three regions for performance assessment ((1) complete tumor: all four tissue classes; (2) tumor core: necrosis, non-enhancing tumor, and enhancing tumor; and (3) enhancing tumor). See Table 1 for quantitative measures and challenge rankings.

tional models were created from the 10 low-grade glioma data sets. Following model construction, weighted importance feature plots described earlier were rendered to provide feedback as to the individual potential predictive accuracy of each feature. The high-grade plots are given in Figure 8 whereas the low-grade plots are given in Figure 9.

Immediately apparent from these plots are the importance of certain features. In general, the features based on the symmetric template are quite discriminative thus justifying the increased computational resources required to generate such features. For single-threaded processing on the cluster, creating the feature images for a single subject required approximately 2 hours of total processing time of which the template registration component took approximately 75%. Additionally, the first order statistical features also seemed fairly important which is something that previous work had demonstrated (e.g., Bauer et al., 2012; Zikic et al., 2012). There are a number of differences, however, between low-grade and high-grade importance outcomes which would also seem to justify the creation of separate glioma class models.

As described earlier, the MICCAI 2013 BRATS challenge data was provided to the competitors in three sets. The Evaluation set was used primarily for training although an initial ranking was performed for all the competitors to aid in determining participation in the workshop. Shortly prior to the actual workshop date, the Leaderboard data was released to the competitors for posting of results and subsequent ranking. Finally, the night before the competition, the Challenge data was made available and used to produce the final competitor ranking. In addition to the Dice overlap measure, additional

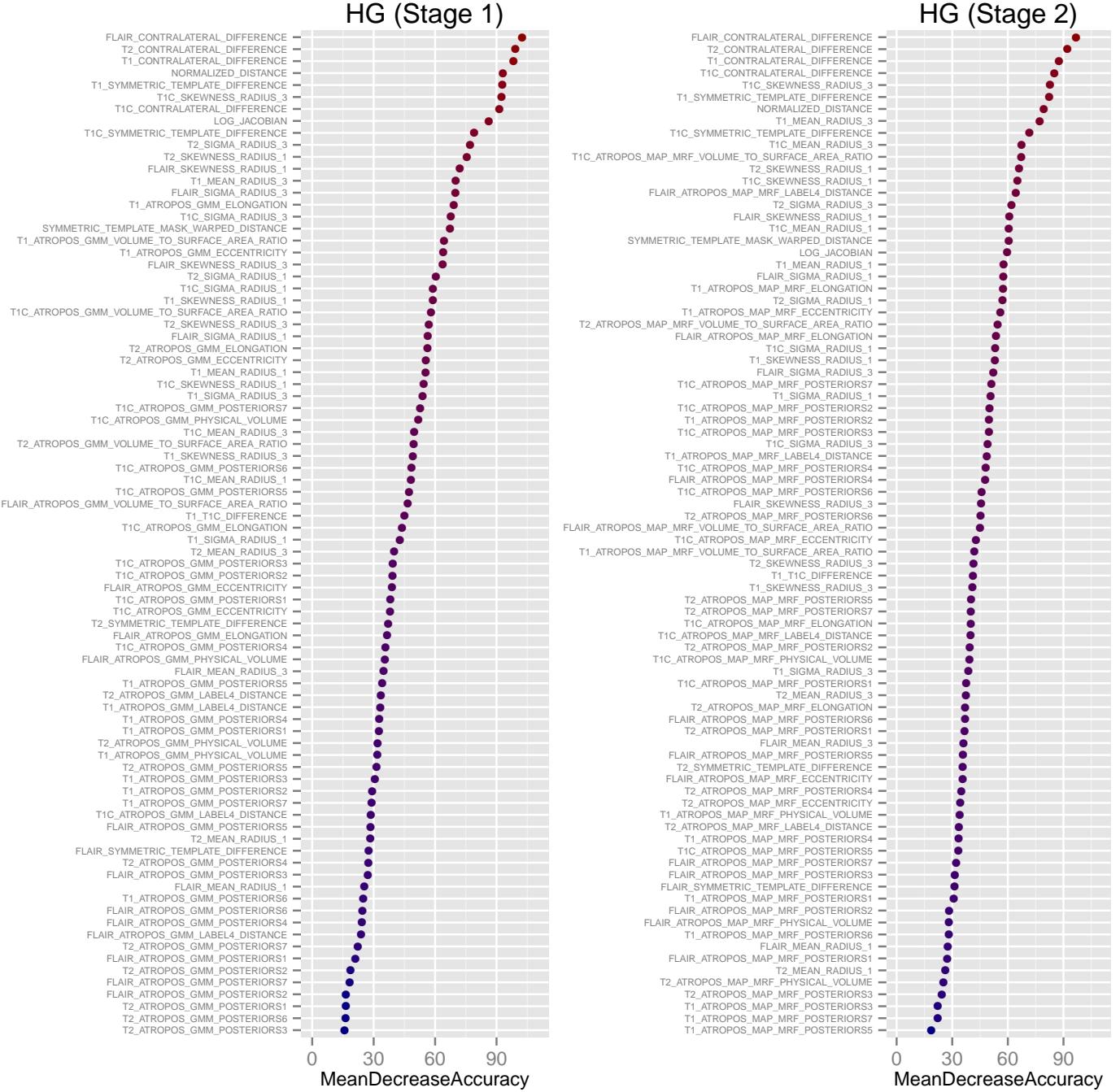


Figure 8: `MeanDecreaseAccuracy` plots generated from the high-grade glioma Stage 1 and Stage 2 RF models. These plots provide a weighted ranking describing the importance of each feature for predictive accuracy.

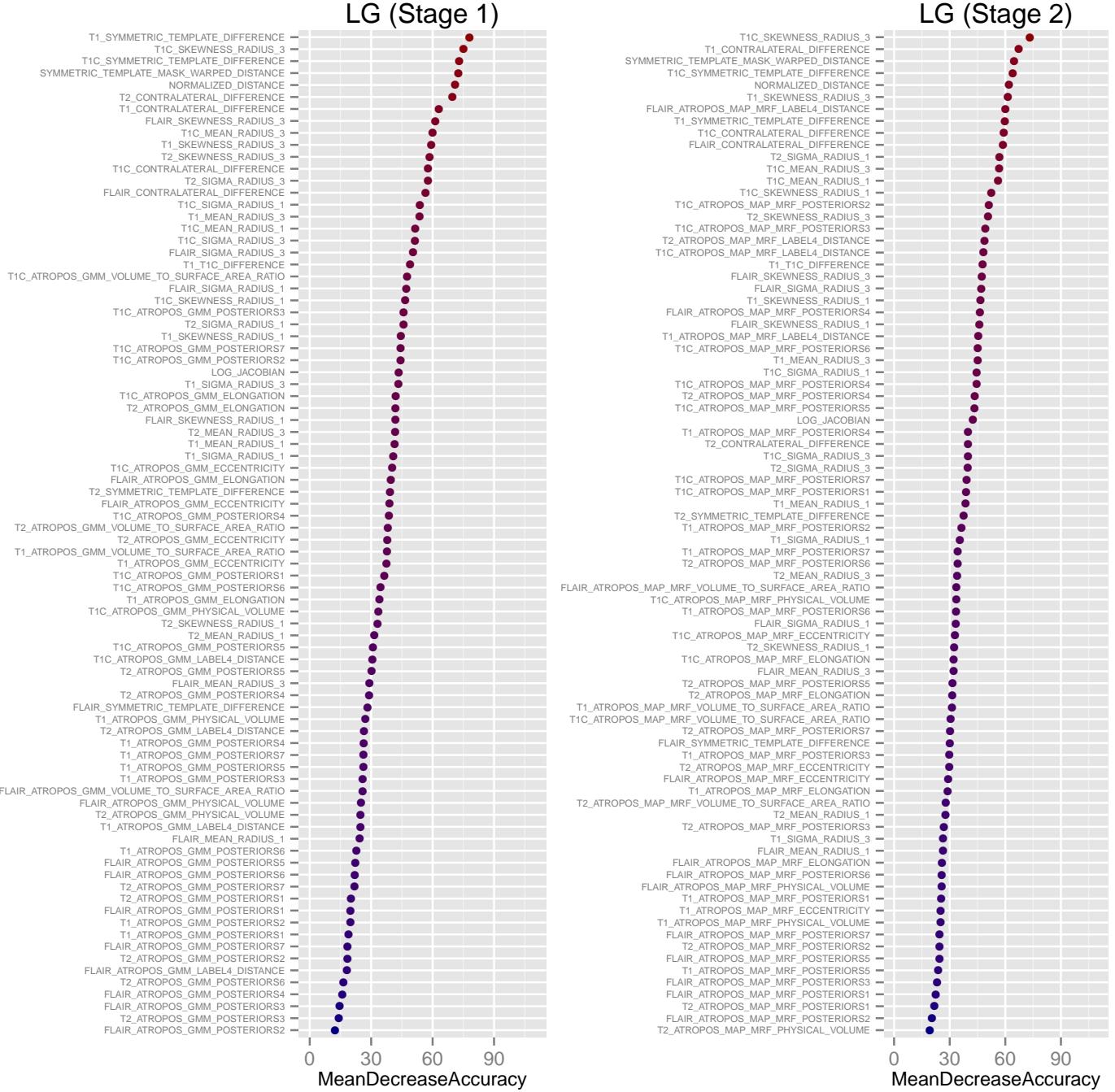


Figure 9: `MeanDecreaseAccuracy` plots generated from the low-grade glioma Stage 1 and Stage 2 RF models. These plots provide a weighted ranking describing the importance of each feature for predictive accuracy.

performance measures for producing the rankings included the positive predictive value, and sensitivity (all of which can be calculated using open source tools such as Tustison and Gee (2009)). For all three data sets, we provide these overlap measures and relative competitor ranking in Table 1. Full competition results can be viewed at <http://www.virtualskeleton.ch>.

4. Discussion and Conclusions

One of the difficulties with the focus of this particular competition is the lack of consensus, even clinically, as to tumor type and extent in the context of medical imaging (Cha, 2005). In fact, current consensus guidelines from the World Health Organization for brain tumor classification are strictly histopathological (Louis et al., 2007) which limit clinical application. Such limitations motivate the use of medical imaging for treatment planning and outcome analyses (Cha, 2005) including more automated methods such as that proposed in this work.

Although previous research has employed RFs for supervised brain segmentation, our contribution of concatenated RF model application and use of symmetric multivariate templates demonstrated good performance in the recent MICCAI 2013 Brain Tumor Segmentation challenge/workshop. In terms of variable importance, the latter provided several highly discriminative features which resulted in the top-performing algorithm of the competition. This confirmed what others have found in that RFs provide an excellent framework for prediction in certain medical image analysis problems. Even with relatively few training subjects relative to input variables, the RF models perform well. In addition, this work highlights the value of a symmetric

multiple modality template for clinical disorders in which asymmetry is a hallmark. Furthermore, our challenge-leading results establish the value of such templates in multiple modality prediction tasks in which there is a small training set with an abundance of multivariate data.

Although these competitions are extremely useful as they provide objective assessment of algorithmic performance given the prevalence of selection bias in reporting results in conventional publication venues, even more important is the availing of the actual algorithmic instantiation (i.e. code) so that others can more easily build upon and utilize what has proven effective. This also provides an opportunity for the users to return constructive feedback to the authors thereby improving the original offering.

However, as explained previously, the brain tumor segmentation methodology that we have made available is only a small part of the larger software package that we have created in *ANTsR*. Not only does *ANTsR* significantly facilitate the development of the work discussed, but it provides an interface to one of the most powerful statistical packages available in R. The combination of the well-known ANTs software package with R provides tremendous potential for future insightful analysis.

References

- Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. *Neural Computation* 9, 1545–1588.
- Angelini, E. D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Dufau, H., 2007. Glioma dynamics and computational models: A review of

- segmentation, registration, and in silico growth algorithms and their clinical applications. *Current Medical Imaging Reviews* 3, 262–276.
- Ashburner, J., Friston, K., Oct 1997. Multimodal image coregistration and partitioning—a unified framework. *Neuroimage* 6 (3), 209–17.
- Avants, B., Duda, J. T., Kim, J., Zhang, H., Pluta, J., Gee, J. C., Whyte, J., Nov 2008a. Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Acad Radiol* 15 (11), 1360–75.
- Avants, B. B., Epstein, C. L., Grossman, M., Gee, J. C., Feb 2008b. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* 12 (1), 26–41.
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., Gee, J. C., Feb 2011a. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage* 54 (3), 2033–44.
- Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., Gee, J. C., Dec 2011b. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics* 9 (4), 381–400.
- Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J. C., Feb 2010. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49 (3), 2457–66.
- Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.-P., Reyes, M., October 2012. Segmentation of brain tumor images based on integrated hierarchical

classification and regularization. In: Proceedings of MICCAI-BRATS 2012. pp. 10–13.

Bauer, S., Nolte, L.-P., Reyes, M., 2011. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Comput Comput Assist Interv* 14 (Pt 3), 354–61.

Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M., Jul 2013. A survey of MRI-based medical image analysis for brain tumor studies. *Phys Med Biol* 58 (13), R97–129.

Breiman, L., 1996. Bagging predictors. *Machine Learning* 24, 123–140.

Breiman, L., 2001. Random forests. In: *Machine Learning*. pp. 5–32.

Cha, S., May 2005. Update on brain tumor imaging. *Curr Neurol Neurosci Rep* 5 (3), 169–77.

Childs, H., Geveci, B., Schroeder, W. J., Meredith, J. S., Moreland, K., Sewell, C., Kuhlen, T., Bethel, E. W., 2013. Research challenges for visualization software. *IEEE Computer* 46 (5), 34–42.

Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K., Jan 2013. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med Image Anal*.

Criminisi, A., Shotton, J., Konukoglu, E., 2011. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Tech. rep., Microsoft Resaerch.

- Das, S. R., Avants, B. B., Grossman, M., Gee, J. C., Apr 2009. Registration based cortical thickness measurement. *Neuroimage* 45 (3), 867–79.
- Durst, C. R., Raghavan, P., Shaffrey, M. E., Schiff, D., Lopes, M. B., Sheehan, J. P., Tustison, N. J., Patrie, J. T., Xin, W., Elias, W. J., Liu, K. C., Helm, G. A., Cupino, A., Wintermark, M., Feb 2014. Multimodal mr imaging model to predict tumor infiltration in patients with gliomas. *Neuroradiology* 56 (2), 107–15.
- Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55, 119–139.
- Geremia, E., Clatz, O., Menze, B. H., Konukoglu, E., Criminisi, A., Ayache, N., Jul 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *Neuroimage* 57 (2), 378–90.
- Geremia, E., Menze, B. H., Ayache, N., October 2012. Spatial decision forests for glioma segmentation in multi-channel MR images. In: Proceedings of MICCAI-BRATS 2012.
- Ho, T. K., 1995. Random decision forests. In: Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 1. pp. 278–282 vol.1.
- Iglesias, J. E., Liu, C.-Y., Thompson, P., Tu, Z., 2010. Agreement-based semi-supervised learning for skull stripping. *Med Image Comput Comput Assist Interv* 13 (Pt 3), 147–54.

- Ince, D. C., Hatton, L., Graham-Cumming, J., Feb 2012. The case for open computer programs. *Nature* 482 (7386), 485–8.
- Kehrer, J., Hauser, H., Mar 2013. Visualization and visual analysis of multifaceted scientific data: a survey. *IEEE Trans Vis Comput Graph* 19 (3), 495–513.
- Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A. D., Bogovic, J. A., Hua, J., Chen, M., Jarso, S., Smith, S. A., Joel, S., Mori, S., Pekar, J. J., Barker, P. B., Prince, J. L., van Zijl, P. C. M., Feb 2011. Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage* 54 (4), 2854–66.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2/3, 18–22.
- Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., Kleihues, P., Aug 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114 (2), 97–109.
- Maurer, C. R., Rensheng, Q., Raghavan, V., 2003. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25 (2), 265–270.
- Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L.,

Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glockner, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2014. The multimodal brain tumor image segmentation benchmark (brats).

URL <http://hal.inria.fr/hal-00935640>

Menze, B. H., Van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P., 2010. A generative model for brain tumor segmentation in multi-modal images. *Med Image Comput Comput Assist Interv* 13 (Pt 2), 151–9.

Ny  l, L. G., Udupa, J. K., Zhang, X., Feb 2000. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 19 (2), 143–50.

Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G., Dec 2003. Automatic brain tumor segmentation by subject specific modification of atlas priors. *Acad Radiol* 10 (12), 1341–8.

Price, S. J., Burnet, N. G., Donovan, T., Green, H. A. L., Pe  a, A., Antoun, N. M., Pickard, J. D., Carpenter, T. A., Gillard, J. H., Jun 2003. Diffusion tensor imaging of brain tumours at 3T: a potential tool for assessing white matter tract invasion? *Clin Radiol* 58 (6), 455–62.

Prima, S., Ourselin, S., Ayache, N., Feb 2002. Computation of the mid-sagittal plane in 3-D brain images. *IEEE Trans Med Imaging* 21 (2), 122–38.

Reynolds, D. A., 2009. Gaussian mixture modeling. In: Li, S. Z., Jain, A. K. (Eds.), *Encyclopedia of Biometrics*. Springer US, pp. 659–663.

Schapire, R., 1990. The strength of weak learnability. *Machine Learning* 5, 197–227.

Tustison, N. J., Avants, B. B., 2013. Explicit B-spline regularization in diffeomorphic image registration. *Front Neuroinform* 7, 39.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., Gee, J. C., Jun 2010. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 29 (6), 1310–20.

Tustison, N. J., Gee, J. C., 2009. Introducing Dice, Jaccard, and other label overlap measures to ITK. *Insight Journal*.

Tustison, N. J., Johnson, H. J., Rohlfing, T., Klein, A., Ghosh, S. S., Ibanez, L., Avants, B. B., 2013. Instrumentation bias in the use and evaluation of scientific software: recommendations for reproducible practices in the computational sciences. *Front Neurosci* 7, 162.

Van Horn, J. D., Toga, A. W., Oct 2013. Human neuroimaging as a “big data” science. *Brain Imaging Behav.*

Verhoek, M., Yaqub, M., McManigle, J., Noble, J. A., 2011. Learning optical flow propagation strategies using random forests for fast segmentation in

- dynamic 2D and 3D echocardiography. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (Eds.), Machine Learning in Medical Imaging. Vol. 7009 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 75–82.
- Viola, P., Jones, M., Snow, D., 2005. Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision 63, 153–161.
- Wilkinson, G. N., Rogers, C. E., 1973. Symbolic description of factorial models for analysis of variance. Journal of the Royal Statistical Society. Series C (Applied Statistics) 22 (3), 392–399.
- Yi, Z., Criminisi, A., Shotton, J., Blake, A., 2009. Discriminative, semantic segmentation of brain tissue in MR images. Med Image Comput Comput Assist Interv 12 (Pt 2), 558–65.
- Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D. H., Demiralp, C., Thomas, O. M., Das, T., Jena, R., Price, S. J., October 2012. Context-sensitive classification forests for segmentation of brain tumor tissues. In: Proceedings of MICCAI-BRATS 2012. pp. 1–9.